J1
.02:S38
8
.2

DEC 1 1986
N.C. DOCUMENTS
N. C. STATE LIBRARY
RALEIGH

SCHS
State Center For Health Statistics

# STATISTICAL PRIMER

Department of Human Resources • Division of Health Services • P.O. Box 2091 • Raleigh, NC 27602-2091 • 919/733-4728

No. 8 October 1986

## THE MEASUREMENT OF ATTRIBUTABLE RISK: A USEFUL TOOL FOR HEALTH ADMINISTRATORS

by

**Kathryn B. Surles and Michael J. Symons†**

Health planners and program managers are often faced with the task of devising some algorithm by which monies will be allocated or a particular prevention/intervention strategy will be focused in order to reduce some unwanted outcome, for example, teenage pregnancy, infant mortality, inadequate prenatal care, or death from a certain cause. The following methodology is not meant to replace the allocation formulae that now exist in the Division of Health Services. Rather, this paper is of an exploratory nature, suggesting that standard quantitative epidemiologic tools may also be used for administrative purposes, albeit that the present approach will not be a panacea for all situations.

Unfavorable rates or large numbers of the unwanted outcome may be considered separately, but inequities may follow either one of these two aspects taken alone. Should counties with large numbers of the unwanted outcome always be funded at the highest levels? Should a small county receive the largest share because it has the worst rate? Should younger and older women necessarily be the focus of infant death interventions because their infants have the higher death rates?

An entity known as "attributable risk" combines the urgency of an unfavorable rate with the frequency of the unwanted outcome. This measure was originally defined by Levin (1,2) and indicates the proportion of the unwanted outcome that is attributable to an "exposure," say living in a particular county, working in a particular occupation, or possessing a certain personal characteristic.

Attributable Risk (AR) is expressed as

$$AR = \frac{e\,(RR-1)}{e\,(RR-1) + 1}$$

where e is the proportion of the entire population that is exposed and RR is the relative risk, i.e., the ratio of the fraction ($f_1$) of those ***exposed*** who experienced the outcome and the fraction ($f_0$) of those ***unexposed*** who experienced the outcome, that is, $RR = f_1/f_0$. By multiplying the result of the AR computation by 100, AR is expressed as a percentage. [Note: The parameter AR has been termed "population attributable risk" and "etiologic fraction" by some authors. We follow Walter (3) and express AR as a percentage.]

Lilienfeld and Lilienfeld (4) provide additional description of attributable risk including standard errors and confidence limits. Walter (3) provides a comprehensive statistical review of this and related epidemiological indices.

The preceding formula is appropriate for prospective data. However, in state health department settings, cross-sectional and retrospective designs are often used. The first two examples that follow involve prospective and cross-sectional data, both yielding estimates of rates. Their use in calculating relative risk is described in Appendices 1 and 2. The third example uses data from a retrospective design for which relative risk is approximated. These details are illustrated in Appendix 3.

### Example 1. Infant Deaths Due to Cause X: A Prospective Design Application (see Appendix 1)

Scenario: A state-level prevention/intervention program wishes to expand its efforts among women whose infants contribute the most to Cause X deaths. The following data for race and maternal age groups represent live-birth cohorts and their corresponding numbers of infant deaths with resulting measures of risk:

---

† A statistical consultant to the SCHS, Dr. Symons is a member of the faculty of the Department of Biostatistics, School of Public Health, University of North Carolina, Chapel Hill.

**Multi-Year Infant Deaths Due to Cause X**

| Race and Maternal Age | Number of Live Births | Number of Deaths | RR* | Percent AR* | Ranks** Number of Deaths | Ranks** RR | Ranks** Pct. AR |
|---|---|---|---|---|---|---|---|
| **Whites** | | | | | | | |
| <18 | 14,332 | 186 | 1.40 | 1.34 | 4 | 4 | 3 |
| 18-34 | 260,307 | 1,808 | 0.52 | -41.97 | 1 | 6 | 6 |
| 35+ | 11,455 | 98 | 0.91 | -0.25 | 5 | 5 | 5 |
| **Nonwhites** | | | | | | | |
| <18 | 16,185 | 236 | 1.59 | 2.21 | 3 | 2 | 2 |
| 18-34 | 114,539 | 1,566 | 1.75 | 17.01 | 2 | 1 | 1 |
| 35+ | 4,923 | 66 | 1.44 | 0.51 | 6 | 3 | 4 |
| Total | 421,741 | 3,960 | - | - | | | |

* See example of computations in Appendix 1.
** Ranked 1-6, high value to low.

From these data, it is seen that being white in combination with being 18-34 is a benefit relative to Cause X infant deaths; AR is high and negative. In contrast, being nonwhite is a detriment, regardless of maternal age. Even in the traditionally low-risk age group 18-34, the AR for nonwhites is substantial at 17%.

Clearly, these data suggest that, in order to reduce the statewide incidence of Cause X infant deaths in the most effective way, nonwhites 18-34 must be targeted. Even though ages under 18 and 35+ are associated with high levels of risk, the numbers of events are small such that improvements in those groups will only marginally reduce the number of Cause X infant deaths statewide.

**Example 2. Cases of Disease X: A Cross-Sectional Design Application (see Appendix 2)**

Scenario: The state has appropriated $500,000 to be distributed among six pilot prevention/intervention projects located in some of the state's higher-risk counties. Data for the six counties are as follows:

**Annual Cases of Disease X**

| Place of Residence | Mid-Year Population | Number of Cases | RR* | Percent AR* | Ranks** Number of Cases | Ranks** RR | Ranks** Pct. AR |
|---|---|---|---|---|---|---|---|
| County A | 1,131 | 140 | 1.28 | 0.13 | 6 | 4 | 6 |
| County B | 1,684 | 202 | 1.25 | 0.16 | 5 | 5 | 5 |
| County C | 9,891 | 1,338 | 1.41 | 1.61 | 2 | 2 | 1 |
| County D | 2,169 | 276 | 1.32 | 0.27 | 4 | 3 | 4 |
| County E | 17,851 | 1,812 | 1.07 | 0.47 | 1 | 6 | 3 |
| County F | 4,432 | 669 | 1.56 | 0.97 | 3 | 1 | 2 |
| A-F Combined | 37,158 | 4,437 | 1.29 | 4.15 | | | |
| All Other Counties | 213,819 | 19,484 | .77 | -23.93 | | | |
| Total | 250,977 | 23,921 | - | - | | | |

* See example of computations in Appendix 2.
** Ranked 1-6, high value to low.

Note that County E would receive the largest slice of the pie based on ***number*** of cases and the smallest slice based on its relative risk. But based on Attributable Risk, County E ranks a distant third among the six counties with only one-half of one percent of the statewide incidence being attributable to living in County E.

Note also that an estimated 4.15% of the statewide incidence is attributable to the six counties combined. In contrast, AR is negative for all other counties combined, meaning that ***on average***, not living in the A–F conglomerate is a relative benefit. It may even be said that the statewide number of Disease X cases may be expected to increase by 23.93% if the protection of living outside the six-county conglomerate were removed.

Now let us say that, in consideration of the state appropriation of $500 thousand, it is decided that each county should receive a base amount of $25 thousand for start-up costs. The balance of $350 thousand will then be allocated proportionate to the county's AR.

Although AR is not additive across the six counties, we assume that it is for the purpose of proportionate allocation. Adding across counties, we obtain a total AR of 3.62%. Each county's percentage contribution to that total is then applied to the $350 thousand and to that amount is added the base amount of $25 thousand as follows:

| | Percent AR | | Money | | |
|---|---|---|---|---|---|
| | Actual | Percent Dist'n. | Allocated via %AR | Base Amount | Total Allocation |
| County A | 0.13 | 3.6 | $ 12,600 | $ 25,000 | $ 37,600 |
| County B | 0.16 | 4.4 | 15,400 | 25,000 | 40,400 |
| County C | 1.61 | 44.5 | 155,750 | 25,000 | 180,750 |
| County D | 0.27 | 7.5 | 26,250 | 25,000 | 51,250 |
| County E | 0.47 | 13.0 | 45,500 | 25,000 | 70,500 |
| County F | 0.98 | 27.0 | 94,500 | 25,000 | 119,500 |
| Total | 3.62 | 100.0 | $350,000 | $150,000 | $500,000 |

Even though County C ranked below County F in relative risk, its number of Disease X cases was twice that in County F. Thus, it gets a larger slice of the pie. County E, on the other hand, had far-and-away the largest number but the least relative risk and therefore gets a much smaller slice than either County C or County F. Even though Counties A and B each get only about 8 percent of the total pie, note that they ranked below average in ***both*** number and risk, relative to the other four counties. Also note that, compared to County C, Counties A and B get only one-fifth to one-fourth as much money, seemingly a mere pittance, but not so little either when one considers that the two counties average only one-eighth as many Disease X cases as County C.

**Example 3. Deaths Due to Cause X in Occupation Group A versus Occupation Group B: A Retrospective Design Application (see Appendix 3)**

Scenario: Two groups of decedents are highly and about equally likely to have died from Cause X and the epidemiologist wishes to select one of the occupation groups for research activities. Data for the two groups of decedents are as follows:

**Annual Deaths in Occupation Groups A and B**

| | Number of Deaths | | | | Ranks | | |
|---|---|---|---|---|---|---|---|
| | Cause X | Other Causes | Odds Ratio* | Pct. AR | Number of Cause X Deaths | Odds Ratio | Pct. AR |
| Occupation A | 785 | 20,156 | 1.61 | 20.0 | 1 | 2 | 1 |
| Occupation B | 304 | 6,575 | 1.65 | 8.1 | 2 | 1 | 2 |
| Total | 1,484 | 49,548 | - | - | | | |

* The Odds Ratio is an estimate of relative risk (RR). See example of computations in Appendix 3.

Clearly, if one had to choose, Occupation A should be selected for research activities related to Cause X . That is because, compared to Group B, identifying and eliminating the etiologic agent(s) responsible for the Group A deaths would be expected to result in a much greater savings in overall deaths due to Cause X.

## DISCUSSION

This seems to us a practical and balanced way of allocating resources, be they monies, manpower, or some other form of supply—scarce medications, for example. In the case of new monies, however, it may be prudent to provide a base amount for start-up costs to all recipients. Had we not established such a threshold amount in Example 2, i.e., had we allocated the entire $500 thousand on the basis of AR, Counties A and B would receive only $18 and $23.5 thousand respectively, both below the $25 thousand deemed necessary to get a project going.

The reader's attention is directed to the appendices which describe the content and calculations of prospective studies (Appendix 1), cross-sectional studies (Appendix 2), and retrospective studies (Appendix 3). Our Example 1 represents the prospective design; a cohort of live births is followed to determine death or survival within the first year of life. Our Example 2 follows the cross-sectional design; individuals from the entire population are dichotomized on the basis of both the outcome (Disease X or not) and the ***concurrent*** exposure (living in a particular county or not). Example 3, on the other hand, represents the retrospective design; "cases" (deaths due to Cause X) are compared with "controls" (deaths not due to Cause X) with respect to their ***previous*** exposure (usual lifetime occupation). In Examples 1 and 2, relative risk (RR) is directly calculable whereas, in Example 3, relative risk is estimated by the Odds Ratio (OR). The reader should also note in Appendix 3 that Cornfield's formula for the Odds Ratio (5) is applicable ***when the unwanted outcome is rare***. The lengthier version is generally applicable and directly follows from Cornfield's derivation.

Lilienfeld and Lilienfeld (4) draw attention to the use of attributable risk in planning disease control programs and in developing strategies for epidemiologic research, and indeed, the simple examples of the present paper point up the usefulness of AR in enabling health administrators to predict the effectiveness of a program in eliminating the etiologic agents underlying a particular exposure. As Walter (3) warns, however, attributable risk should "in no way be regarded as a substitute for relative risk, but rather as an alternative or additional dimension of health hazard appraisal." The identification of a particular exposure with a high relative risk may provide important clues as to the disease mechanism. However, if the exposure is rare in the population, attributable risk will be low and the exposure of little interest to health administrators less concerned with etiology than with reducing the incidence of the unwanted outcome in the population as a whole. Walter concludes that a rational choice between alternative programs, say between a lifestyle intervention and some environmental modification, "...must be based at least in part on an assessment of their ***potential*** benefits; it appears that attributable risk is a suitable measure in this context."

A final word is that use of AR may sometimes be circumvented by limitations on the per-individual cost of "treatment," meaning either an intervention or a research activity. For example, if we have resources enough to "treat" only 200 out of a possible 2,000 individuals, then naturally those at the highest ***risk*** should be selected. The whole point is, what approach affords us the greatest return on our resources? The answer to this question becomes increasingly important as we attempt to achieve the most efficient and equitable use of public health's diminishing dollar.

# APPENDIX 1

## Relative Risk and Attributable Risk in Prospective Studies

Relative Risk (RR) is a calculation that is naturally made with prospective data. If two groups are identified as exposed or not, i.e., having or not having a particular characteristic or exposure, and are followed to see how many develop an (unwanted) outcome within the time frame of the study, the study is a prospective one. The data take the following form:

| | | Unwanted Outcome | | |
|---|---|---|---|---|
| | | **Yes** | **No** | **Total** |
| | Yes | a | b | u |
| **Exposure** | No | c | d | v |
| | Total | r | s | N |

and $RR = \frac{a/u}{c/v} = f_1 / f_0$. If the total N is a sample of the entire population, then e = u/N; otherwise e must be estimated independently. Then

$$AR = \frac{e\,(RR-1)}{e\,(RR-1) + 1}$$

which, when multiplied by 100, is the percent attributable risk associated with the exposure. Data from other designs require different approximations to relative risk as described in Appendices 2 and 3.

The data table and computations for Whites under 18 in Example 1 are as follows:

| | | Outcome | | |
|---|---|---|---|---|
| | | **Death** | **Survival** | **Total** |
| | Whites <18 | 186 | 14,146 | 14,332 |
| **Exposure** | All Others | 3,774 | 403,635 | 407,409 |
| | Total | 3,960 | 417,781 | 421,741 |

$$RR = \frac{186/14{,}332}{3{,}774/407{,}409} = 1.40$$

$$e = 14{,}332/421{,}741 = .034$$

$$AR = \frac{.034(1.40-1)}{.034(1.40-1)+1} = .0134 \text{ or } 1.34\%$$

# APPENDIX 2

## Relative Risk and Attributable Risk in Cross-Sectional Studies

If two groups are identified that differ in both outcome (e.g., having or not having a disease) and concurrent exposure (e.g., living in a particular county or not), the study is "cross-sectional." For these computations, data take the same form as described in Appendix 1, and the fraction diseased within time t with rate m operative is approximated by the exponential model, $f = 1-\exp(-mt)$. Then

$$RR = \frac{f_1}{f_0} = \frac{1-\exp(-m_1 t)}{1-\exp(-m_0 t)},$$

where t is the length of the period of study, $m_1$ is the estimated rate in the exposed group; and $m_0$ is the estimated rate in the comparison group. With $e = u/N$, again

$$AR = \frac{e(RR-1)}{e(RR-1)+1}$$

The data and computations for County A in Example 2 are as follows:

| | | Outcome | | |
|---|---|---|---|---|
| | | **Disease X** | **Not Disease X** | **Mid-year Population** |
| **Exposure** | County A | 140 | 991 | 1,131 |
| | Not County A | 23,781 | 226,065 | 249,846 |
| | Total | 23,921 | 227,056 | 250,977 |

$m_1 = 140/1{,}131$, $m_0 = 23{,}781/249{,}846$, $t = 1$ (year), and

$$RR = \frac{1-\exp(-0.1238)}{1-\exp(-0.0952)} = \frac{1-.8836}{1-.9092} = 1.28$$

Now, $e = 1{,}131/250{,}977 = .0045$ so that

$$AR = \frac{(.0045)(1.28-1)}{(.0045)(1.28-1)+1} = .0013 \text{ or } .13\%$$

When the rates are small, the approximation of relative risk is simply the ratio of the rates, i.e., $RR = m_1 / m_0$, following a linearized Taylor series approximation of the exponential function. Here, $\frac{140/1131}{23{,}787/249{,}846} = 1.30$, quite close to 1.28.

# APPENDIX 3

## The Odds Ratio and Attributable Risk in Retrospective Studies

### Part I. Description and Calculations

If two groups are identified that differ only in whether or not they had the unwanted outcome, the study is a "retrospective" one; the two groups are examined to see how many ***previously*** had the exposure. Since incidence rates are not determined in a retrospective study, the relative risk (RR) is estimated by the odds ratio (OR) which is computed as:

$$OR = \frac{p_1}{1-p_1} \Big/ \frac{p_2 + p(p_1 - p_2)}{1-p_2 - p(p_1 - p_2)},$$

where $p_1$ is the fraction exposed among those with the undesirable outcome, $p_2$ is the fraction exposed among those with all other outcomes, and p is the chance of the unwanted outcome in the total population of the exposed and unexposed.

For these computations, data take the same form as described in Appendix 1. The data table and computations for Occupation Group A in Example 3 are as follows:

| | | Outcome | | |
|---|---|---|---|---|
| | | **Death from Cause X** | **Death Not from Cause X** | **Total** |
| **Occupation** | Group A | 785 | 20,156 | 20,941 |
| | Not Group A | 699 | 29,392 | 30,091 |
| | Total | 1,484 | 49,548 | 51,032 |

$p_1 = a/r = 785/1,484 = .5290$

$p_2 = b/s = 20,156/49,548 = .4068$

$p = r/N = 1,484/51,032 = .0291$

$$OR = \frac{.5290}{1-.5290} \Big/ \frac{.4068 + .0291(.5290 - .4068)}{1-.4068 - .0291(.5290 - .4068)} = 1.61$$

$e = 20,941/51,032 = .41035$

$$AR = \frac{(.41035)(1.61-1)}{(.41035)(1.61-1) + 1} = .2002 \text{ or } 20.02\%$$

To simplify things, the reader will note that, when the outcome is rare, p is small and the formula for OR reduces to Cornfield's famous Odds Ratio (5):

$$OR = \frac{p_1}{(1-p_1)} \Big/ \frac{p_2}{(1-p_2)}$$

This is equivalent to OR = ad/bc, giving the following results for the example used above:

$$OR = (785)(29,392)/(20,156)(699) = 1.64$$

$$AR = \frac{(.41035)(1.64-1)}{(.41035)(1.64-1) + 1} = .2080 \text{ or } 20.8\%$$

In this case, the simple computation ad/bc measures OR with relatively little error, but users should keep in mind that the error can be substantial for more common events.

**Part II. Comparisions to the Proportionate Mortality Ratio (PMR)**

In recent years, various researchers at the state and national levels have become involved in the use of death certificates to investigate cause-specific mortality by occupation and industry of the decedent (decedent's usual lifetime job). Those efforts have generally used the PMR method of risk appraisal, where PMR is the ratio of the observed number of cause-specific deaths in the occupation group to the expected number obtained by multiplying the proportion of cause-specific deaths in a reference population by the total number of deaths in the occupation group. While we concur in the use of that method as one alternative, we believe that OR and the subsequent AR provide additional important information. Consider the following statistics from Example 3 of the present paper:

| | PMR | OR | %AR |
|---|---|---|---|
| Occupation A | 1.29 | 1.61 | 20.0 |
| Occupation B | 1.52 | 1.65 | 8.1 |

Clearly, if one had to choose between the two groups for research or intervention activities, Occupation B would be the better choice on the basis of PMR while Occupation A would win handily on the basis of AR. When one compares the relative risk in one group to that in the other via OR as an estimate of RR, we have virtually a toss of the coin.

Miettinen and Wang (6) previously expressed a preference for the OR over the PMR. Their arguments were based upon numerical comparisons of the two statistics as well as epidemiologic interpretations.

STATE LIBRARY OF NORTH CAROLINA
3 3091 00004 2739

## REFERENCES


1. Levin, M. L., "The Occurrence of Lung Cancer in Man," *Acta Unio. Internat. Contra Cancrum* 9, 1953, pp. 531-541.

2. Levin, M.L., and Bertnel, R., "Simple Estimation of Population Attributable Risk from Case-Control Studies," *American Journal of Epidemiology* 108, 1978, pp. 78-79.

3. Walter, S.D., "The Estimation and Interpretation of Attributable Risk in Health Research," *Biometrics* 32, December 1976, pp. 829-849.

4. Lilienfeld, A.M., and Lilienfeld, D.E., *Foundations of Epidemiology*, 2d. ed., Oxford University Press, New York, 1980, pp. 191-225 and 342-347.

5. Cornfield, J., "A Method ot Estimating Comparative Rates from Clinical Data: Applications to Cancer of the Lung, Breast, and Cervix," *Journal of the National Cancer Institute* 14, 1951, pp. 1269-1275.

6. Miettinen, O.S. and Wang, J.D., "An Alternative to the Proportionate Mortality Ratio," *American Journal of Epidemiology* 114, 1981, pp. 144-148.

Department of Human Resources
Division of Health Services
State Center for Health Statistics
P.O. Box 2091
Raleigh, N.C. 27602-2091
919/733-4728

BULK RATE
U.S. Postage PAID
Raleigh, N.C. 27602-2091
Permit No. 1862

CHERYL W. MCLEAN
ASSISTANT DOCUMENTS LIBRARIAN
DIVISION OF STATE LIBRARY
ARCHIVES/LIBRARY BUILDING
RALEIGH, NC INTEROFFICE

400 copies of this public document were printed at a cost of $60.00 or 15¢ per copy.